

---

# Accurate Estimation of Correlation Energies Using Locally Dense Basis Sets

---

D. B. CHESNUT\* and E. F. C. BYRD

*P.M. Gross Chemical Laboratory, Duke University, Durham, North Carolina 27708*

*Received 17 August 1995; accepted 6 November 1995*

---

## ABSTRACT

Locally dense basis sets using the mixed 6-311G(*d*, *p*)/3-21G basis can be used to reproduce total energies and correlation energies after empirical adjustment to 2–4 kcal/mol for a variety of small and medium size molecules containing hydrogen, carbon, and oxygen. Post-Hartree–Fock methods can be calculated faster by this method by factors of 2–3, in general, and higher in the presence of high molecular symmetry; density functional approaches take longer and are impractical in the locally dense basis set approach. It is shown that the correlation energy in two of the better characterized density functional approaches is generally significantly larger than that of the post-Hartree–Fock treatments studied here and appears to be insensitive to the basis set employed. © 1996 by John Wiley & Sons, Inc.

---

## Introduction

While our ability to study increasingly more complex and interesting chemical systems improves, there still exist temporal and spatial limits to the size of system that can be studied. Accordingly, approximations which allow one to look at specific properties in an accurate manner continue to be useful. In the present paper we report on the use of locally dense basis sets<sup>1,2</sup> and what appears to be a general additivity relationship that allows one to estimate rather accurately correlation energies in complex systems.

Some of the standard methods for including correlation to varying degrees include many-body

or Møller–Plesset perturbation theory,<sup>3–6</sup> limited configuration interaction such as coupled cluster theory,<sup>7–12</sup> quadratic configuration interaction,<sup>13</sup> and the use of Brueckner orbitals<sup>14,15</sup> and, more recently, density functional theory.<sup>16</sup> While each method is successful to some degree, the standard quantum mechanical approaches are very expensive in terms of computer time and space, and are sensitive to the number of basis functions employed. While Hartree–Fock theory formally scales as  $N^4$  (where  $N$  is the number of basis functions), the standard quantum mechanical correlation treatments scale to even higher powers. Møller–Plesset theory scales approximately as  $N^5$  in second order (MP2) and as  $N^6$  in third order (MP3),<sup>17</sup> and to obtain accurate results, large basis sets are generally required, which compounds the problem. Density functional methods are gaining

\* Author to whom all correspondence should be addressed.

widespread use because of their remarkable relative computational economy, but the correct energy functional is not yet known, and the approach in this area, more often than not, is semiempirical.

Methods which yield adequate correlation energies with smaller basis sets allow more rigorous treatments of smaller molecules and permit calculations on larger systems presently inaccessible to post-Hartree-Fock methods. For example, the additivity of correlation and basis set enhancement effects in obtaining *relative* energies ( $\Delta E$  values) has been investigated<sup>18-20</sup> in which one assumes that the effect of the addition of polarization functions to the basis set is the same at the HF (Hartree-Fock) and MP3 levels [eq. (1)] or where one assumes that the addition of such functions is the same at the MP2 and MP3 levels [eq. (2)]. This approach has been used in Pople's G1,<sup>21,22</sup> G2,<sup>23</sup>

$$\begin{aligned} \Delta E(\text{MP3}/6\text{-}31\text{G}^{**})_{\text{HF}, 6\text{-}31\text{G}} \\ = \Delta E(\text{MP3}/6\text{-}31\text{G}) \\ + \Delta E(\text{HF}/6\text{-}31\text{G}^{**}) \\ - \Delta E(\text{HF}/6\text{-}31\text{G}) \end{aligned} \quad (1)$$

$$\begin{aligned} \Delta E(\text{MP3}/6\text{-}31\text{G}^{**})_{\text{MP2}, 6\text{-}31\text{G}} \\ = \Delta E(\text{MP3}/6\text{-}31\text{G}) \\ + \Delta E(\text{MP2}/6\text{-}31\text{G}^{**}) \\ - \Delta E(\text{MP2}/6\text{-}31\text{G}) \end{aligned} \quad (2)$$

and G2(MP2)<sup>24</sup> approaches, and has been shown to be very useful and quite accurate.<sup>25</sup>

Pulay<sup>26</sup> has been chiefly responsible for what is referred to as the local treatment of correlation in which the inclusion of correlation effects from neighboring and strongly localized molecular orbitals is emphasized. This approach results in significant computation savings even for molecules of relatively modest size, but has the drawback of the loss of simplicity of the wave function and the necessarily arbitrary nature of the neglected contributions. Dykstra and Kirtman<sup>27</sup> have reviewed a local space approximation which focuses on changes in electronic structure caused by spatially constrained chemical modifications to a system.

We first employed locally dense basis sets in chemical shielding calculations.<sup>1,2</sup> Chemical shielding is an area where, for the most part, the Hartree-Fock approach has been very successful. However, here, too, there are situations where correlation effects are important and where Hartree-Fock results are in significant error. Examples of including correlation in shielding calculations have recently been reviewed.<sup>28,29</sup> As Saebo

and Pulay point out,<sup>26</sup> the empirical transferability of the ground-state properties of molecules testifies to the localized nature of such systems. We have used this philosophy in the Hartree-Fock calculation of NMR chemical shielding. The idea is to use a relatively large (locally dense) set of functions on the resonant atom of interest and considerably smaller (attenuated) sets of functions on other atoms in the molecule. In our initial study<sup>1</sup> it was suggested that for atoms in the first long row of the Periodic Table a dense basis of 6-311G(*d*, *p*)<sup>30,31</sup> and an attenuated set 3-21G<sup>32</sup> worked well together. A number of applications of this technique have been reported in the literature since then involving the amide group in glycylglycine<sup>33</sup> and its dihydrate,<sup>34</sup> the proton anisotropy in (H<sub>2</sub>O)<sub>17</sub> as a model for ice,<sup>35</sup> and the calculation of chemical shielding in model protein systems.<sup>36</sup> While first efforts were focused on the idea of a single atom being dense, later applications<sup>2,33-36</sup> showed that one can rather select a multiatom segment or chemical functional group (the "NMR chromophore") to be dense in order to obtain considerably improved results. The local behavior of many phenomena in chemistry is well known, and this appears to be the reason for the success of the locally dense basis set approach as employed in the Hartree-Fock calculations of NMR shieldings.

Interestingly enough, the Hartree-Fock energies found in our NMR shielding calculations using the locally dense basis set approach showed a basic additivity.<sup>2</sup> When we considered the difference in total energy between a balanced (not locally dense) large and a balanced small basis set calculation, we found that this difference could be represented rather accurately by a summation of differences between the locally dense calculations and their completely attenuated counterparts. That is, if we designate by capital letters a dense basis in a two-chromophore (two-fragment) molecule (for simplicity) and the attenuated basis by lowercase letters, we found the relationship

$$\begin{aligned} E(AB) - E(ab) \cong (E(Ab) - E(ab)) \\ + (E(aB) - E(ab)) \end{aligned} \quad (3)$$

held remarkably well.

In the present work we show that more general energy differences between the large 6-311G(*d*, *p*) and small 3-21G basis sets as well as correlation energy differences can be calculated with the locally dense basis set approach in just the same way that the Hartree-Fock energies are reproduced by

eq. (3). The approach is applied to a variety of small and medium size molecules containing hydrogen, carbon, and oxygen using a variety of post-Hartree-Fock and density functional methods. Applying an empirical correction dependent on the number of electrons involved, the locally dense method reproduces energy differences to better than 4 kcal/mol on average and correlation energy differences to better than 2 kcal/mol, results that clearly are near "chemical accuracy."

## Details of the Calculations

Locally dense basis sets are defined by using large (dense) sets of functions on one fragment of a molecule while using smaller (attenuated) sets of functions elsewhere. The molecular fragment for which the locally dense basis functions are employed is arbitrary, but past experience with chemical shielding<sup>2</sup> and the results discussed here indicate that the use of chemically meaningful fragments yields the best results. In some cases this fragment can be as small as a single atom, while in others a chemical functional group is appropriate. The fragmentation used in the present calculations is described below.

The *presumption* behind the use of locally dense basis sets for the calculation of energies is that the difference in energy for a molecule using basis sets of two differing sizes is approximately equal to a sum of the energy differences of the molecular fragments comparing the locally dense basis set energy to that of the smaller basis set. This result is illustrated in eq. (4), where  $d$  stands for the locally dense basis and  $a$  stands for the smaller or attenuated basis on the various fragments  $F_i$  of the molecule employing method  $M$ :

$$\begin{aligned} E^{(M)}(F_j = d, \text{all } j) \\ = E^{(M)}(F_j = a, \text{all } j) \\ + \sum_k (E^{(M)}(F_k = d, F_j = a, j \neq k) \\ - E^{(M)}(F_j = a, \text{all } j)) \end{aligned} \quad (4)$$

We presume that method  $M$  can be any particular quantum mechanical approach including post-Hartree-Fock or density functional methods as well as the self-consistent Hartree-Fock (HF) theory itself. As usual we define the correlation energy for method  $M$  in basis set  $s$  [either the balanced dense ( $d$ ) or balanced attenuated ( $a$ ) set] as the difference between the energy for method

$M$  and that of the Hartree-Fock approach as in eq. (5a), and we further define a correlation energy for a locally dense basis set approach as in eq. (5b). Because eq. (4) is presumed to hold for the Hartree-Fock method

$$\begin{aligned} E_{\text{corr}}^{(M)}(F_j = s, \text{all } j) \\ = E^{(M)}(F_j = s, \text{all } j) - E^{(\text{HF})}(F_j = s, \text{all } j) \quad (5a) \\ E_{\text{corr}}^{(M)}(F_k = d, F_j = a, j \neq k) \\ = E^{(M)}(F_k = d, F_j = a, j \neq k) \\ - E^{(\text{HF})}(F_k = d, F_j = a, j \neq k) \end{aligned} \quad (5b)$$

itself, we can readily derive the expression shown in eq. (6) which relates the correlation energy itself between large and small balanced basis sets to a sum over the fragment correlation energies of the locally dense basis set approach:

$$\begin{aligned} E_{\text{corr}}^{(M)}(F_j = d, \text{all } j) \\ = E_{\text{corr}}^{(M)}(F_j = a, \text{all } j) \\ + \sum_k (E_{\text{corr}}^{(M)}(F_k = d, F_j = a, j \neq k) \\ - E_{\text{corr}}^{(M)}(F_j = a, \text{all } j)) \end{aligned} \quad (6)$$

Although eq. (4) is used to derive eq. (6), since eq. (4) is an approximation, we have no a priori reason to believe one will be more accurate than the other. We shall find that the determination of the correlation energies as given in eq. (6) is somewhat more accurate than the energy differences for method  $M$  as described by eq. (4).

The utility of the method lies in the fact that in the post-Hartree-Fock methods the small or attenuated basis set calculations and those of the locally dense fragments require considerably less time and memory than a fully balanced calculation using the larger basis set. Because some of these methods scale as very large powers of the molecular size, considerable savings can be realized employing this approach. Such is *not* the case for the density functional methods we investigated which, as we will show, are normally *more* expensive in the locally dense approach compared to the conventional balanced large basis set approach. The success of the method in both cases, however, points to a localized nature of the energy contributions of fragments to the total energy and the correlation energy and to a relative insensitivity of the description of the overall molecular wave

function outside the locale of the locally dense fragment.

In the present calculations we have restricted our considerations to the mixing of only two basis sets in the locally dense basis set approach, although this is not necessary and, indeed, in applications of this type of technique independently implemented by Huber and co-workers<sup>37-39</sup> and others,<sup>2</sup> more than two basis sets are employed. Here we use as the locally dense or large basis set the 6-311G(*d*, *p*) basis of Krishnan et al.<sup>30,31</sup> and for the smaller basis the 3-21G split valence basis of Binkley, Pople and Hehre.<sup>32</sup> The latter is a relatively small basis which includes no polarization functions, but has some flexibility in the valence shell, while the larger 6-311G(*d*, *p*) basis is a relatively large basis (for calculations on other than very small molecules) parameterized from MP2 atomic calculations. For our calculations we employed 6 Cartesian *d*-functions and, in a few calculations, 10 *f*-functions. While the 6-311G basis has been shown to be better represented as 63-11G in the *s*-space,<sup>40,41</sup> the use of six *d*-functions provides an additional 3*s*-type Gaussian that will tend to restore the triple-valence-split nature of that basis. All of our calculations were carried out employing the Gaussian94 code<sup>42</sup> on a Gray Y-MP computer located in the North Carolina Supercomputing Center.

The methods involved in the present study include Hartree-Fock theory (HF), second order many-body perturbation theory (second order Møller-Plesset theory, MP2<sup>3-6</sup>), coupled clusters and quadratic configuration interaction theories with singles, doubles, and noniterative triples [CCSD(T)<sup>12</sup> and QCISD(T)<sup>13</sup>, respectively], and Brueckner doubles with noniterative triples (BD(T)<sup>15</sup>) using the frozen core approximation. Post-Hartree-Fock methods have recently been reviewed.<sup>43,44</sup> Two gradient-including density functional methods were employed; these were Becke's three component approach<sup>45</sup> to the exchange-correlation functional (B3PW91) and that density functional approach (BLYP) which employs Becke's gradient-including exchange function<sup>46</sup> with the gradient-including correlation functional of Lee, Yang, and Parr.<sup>47</sup>

The example of low order Møller-Plesset theory (MP2) was chosen for its applicability to relatively large chemical systems. The other post-Hartree-Fock approaches are much more sophisticated and also much more CPU expense. The CCSD(T), QCISD(T), and BD(T) methods first differ in fifth order Møller-Plesset theory and formally scale

as  $N^6$  (with one  $N^7$  noniterative triples calculation).<sup>15</sup> The Brueckner doubles approach has been characterized as more "elegant" than its competitors at this level,<sup>15</sup> but while the equations determining the excitations from the Hartree-Fock ground state are simpler, the method suffers in that it is more difficult to determine the Brueckner orbitals than the canonical Hartree-Fock orbitals. The BLYP density functional approach has generally exhibited good results for both energies and structures,<sup>48-51</sup> while Becke's three parameter approach<sup>45</sup> is a good example of density functional theory optimized in terms of a set of semiempirical coefficients determined by an appropriate fit of experimental data.

The molecules involved in the present study involve only hydrogen, carbon, and oxygen and are indicated in the data tables (Tables II and IV). The molecules include acetylene, ethylene, ethane, ethanol, allene, propyne, cyclopropane, propene, acetaldehyde, propane, ethanol, trans butadiene, cyclobutane, acetone, acetic acid, furan, benzene, and cyclohexane. The set of molecules represents both cyclic and noncyclic structures, single and multiple bonds, conjugated and nonconjugated systems, and molecules with lone pairs. All the structures were optimized at the MP2 level with the large balanced 6-311G(*d*, *p*) basis. The fragmentation employed basically consists of individual singly bonded carbons and any attached hydrogens, OH groups, *both* atoms and their attached hydrogens for multiply bonded species such as the —CHCH<sub>2</sub> group in propene, the HCCH group in furan, the —CCH group in propyne, the CHO and CO groups in acetaldehyde and acetone, respectively, and the CO group in acetic acid. In benzene two adjacent carbons and their attached hydrogens were taken as a single fragment.

## Results and Discussion

### MOLECULAR GEOMETRIES

The present studies involving locally dense basis sets employed geometries optimized at the MP2 level. We also performed optimizations at the HF, BLYP, and B3PW91 approaches in order to compare these four methods. The MP2 approach was the best, with both density functional treatments close behind; all three methods were superior to the Hartree-Fock approach. For the 17 molecules studied for which experimental data

was available,<sup>52</sup> bond length root-mean-square errors (rmse) for the MP2, B3PW91, BLYP, and HF methods were 0.0075, 0.0086, 0.0098, and 0.0172 Å, respectively. The rmse values for angles were 0.78, 0.95, 0.99, 1.03°, with the same ordering of methods.

### CALCULATION TIMES AND EFFICIENCY OF THE METHOD

The central aspects of the use of locally dense basis sets in the calculation of energies has to do first with the times involved and, second, with an understanding of why this method works. Here we discuss the efficiency of the method in terms of calculation CPU times on a Cray Y-MP 8/464 computer located in the North Carolina Supercomputing Center. Both relative and absolute times are important. While relative times can be expected to carry over approximately from one platform to another, real calculations are carried out in real time and require absolute CPU allocations. The advantage of using locally dense basis sets in the calculation of Hartree-Fock and, more importantly, the Post-Hartree-Fock methods is the large exponential dependence upon system size of these approaches. While several calculations are required to generate the total molecular energy in the locally dense basis set approach, these many calculations generally will require less time than that of a single calculation with a fully balanced large basis set. Furthermore, the locally dense basis set approach requires no modification to the system's Hamiltonian or the involvement of other nonstandard techniques; one simply uses a mixed basis set with standard computational codes.

Table I shows some times (in minutes) and efficiencies for molecules in the present study in the QCISD(T) and BLYP approaches. The efficiencies listed in the table represent the ratio of the time necessary to carry out a balanced calculation with the large [6-311G(*d*, *p*)] basis divided by the sum of the times required to carry out the various fragment calculations in the locally dense basis set approach plus the small (3-21G) balanced basis set calculation. Acetic acid and furan are fairly representative of the results in this study in that a number of fragment calculations are required and relatively little use of symmetry (in reducing the number of calculations) is used (none in acetic acid). Benzene and especially cyclohexane benefit greatly in this regard by high symmetry. Cyclohexane is a best possible case in that only one fragment (the CH<sub>2</sub> fragment) needs to be calculated.

**TABLE I.**  
Calculation Times *t* (in minutes), and Efficiencies  $\epsilon$  for Four Molecules in the Present Study in the QCISD(T) and BLYP Approaches.<sup>a</sup>

	QCISD(T)		BLYP		<i>r</i>
	<i>t</i>	$\epsilon$	<i>t</i>	$\epsilon$	
Acetic acid	53.6	1.71	1.9	0.34	28
Furan	95.3	2.27	2.3	0.50	41
Benzene	343.6	5.29	2.1	0.54	164
Cyclohexane	1242.6	8.94	3.6	0.54	345

<sup>a</sup> The times are for balanced calculations using the 6-311G(*d*, *p*) basis set, while the efficiency is defined as the time required for the balanced 6-311G(*d*, *p*) calculation divided by the sum of the times required for the 6-311G(*d*, *p*) / 3-21G locally dense and 3-21G balanced calculations. Also shown is the ratio *r* of times for the post-Hartree-Fock QCISD(T) and BLYP density functional calculations.

Generally, one can expect an efficiency enhancement by a factor of 2–3 using locally dense basis sets, with systems of higher symmetry (like benzene and cyclohexane) showing much larger efficiencies.

The data shown for the BLYP density functional method show that the application of the locally dense basis set approach with this technique (and the B3PW91 approach) is literally a waste of time. It generally takes more time (by roughly a factor of 2) to employ the locally dense basis set approach with the density functional methods. We note also the well known large difference of calculational times comparing the high level QCISD(T) method and the BLYP density functional approach. Cyclohexane requires nearly 21 hours of CPU time to complete while BLYP requires less than 4 minutes! One's enthusiasm for density functional techniques and their vast superiority in time of calculation is modified only by the fact that we are still not sure what the correct equations to use are. We know that we are "right" in the application of the post-Hartree-Fock methods like QCISD(T) in that we know the base from which the method springs and the conditions under which it is applicable; such is not yet the case with density functional methods since the exact functional for the energy is as yet unknown. However, the fact that one can carry out a BLYP or B3PW91 calculation of cyclohexane over 300 times faster than a sophisticated post-Hartree-Fock calculation illustrates one of the reasons why research in the density functional area is presently so intense.

The loss of symmetry which the use of locally dense basis sets can effect is, of course, a flaw of

the method. This is undoubtedly the reason why the density functional theory calculations take longer in the locally dense basis set approach than do the balanced calculations. It obviously affects such systems as benzene whose  $D_{6h}$  symmetry is reduced to  $C_{2v}$  with the introduction of the locally dense HCCH fragment; even so, the savings in time in most cases with the use of post-Hartree-Fock methods offsets this symmetry loss effect. It is remarkable that, as we shall see, the energies calculated in the symmetry reduced fragmented systems can still be summed to reproduce the balanced energies (and correlation energies) of the systems where symmetry is maintained (the balanced basis set calculations).

### LOCALLY DENSE ENERGY AND CORRELATION ENERGY DIFFERENCES

The energy differences and correlation energy differences between the locally dense approach and the balanced large basis set approach are tabulated in Table II. The results for the CCSD(T) and BD(T) methods were virtually identical to the QCISD(T) approach, so we tabulated data only for the latter method; the average spread of results for these three post-Hartree-Fock methods was only 0.03 kcal/mol. It was found best to maintain dou-

ble and triple bond fragments intact in the larger molecule such as propene, propyne, trans butadiene, and benzene; complete fragmentation in these larger molecules increases the error by about a factor of 3. (Accordingly, while the locally dense data for both ethylene and acetylene are included in Table II, they are not plotted in the figures.) We note that these larger molecules are also characterized by symmetry which leads to orbital degeneracy near the highest occupied molecular orbital-lowest unoccupied molecular orbital (HOMO-LUMO) gap. As one might expect, the use of a locally dense basis set which tends to destroy such symmetry should be less likely to work in such cases.

There are several other cases which bear mentioning. Allene ( $H_2CCCH_2$ ) has no appropriate way to be fragmented and maintain the double bond entries. Accordingly, it is not plotted or included in Table II nor included in the regression fits. Cyclopropane and cyclobutane—both strained systems—simply fail when totally fragmented into  $CH_2$  fragments, their energies being again about a factor of 3 higher than what might be expected for molecules with the same number of electrons. Orbital degeneracy and the fact that strain will tend to destabilize these orbitals are likely factors in the difficulties these two molecules present.

TABLE II.

Differences between Locally Dense and Balanced Energies (energy differences, first five columns) and Locally Dense and Balanced Correlation Energies (correlation energy differences, last four columns).<sup>a</sup>

	Energy Differences					Correlation Energy Differences			
	HF	MP2	QCISD(T)	BLYP	B3PW91	MP2	QCISD(T)	BLYP	B2PW91
$C_2H_2^b$	-14.74	-17.85	-18.91	-10.90	-11.57	-3.11	-4.18	3.83	3.16
$C_2H_4^b$	-7.32	-10.80	-11.64	-4.58	-4.95	-3.48	-4.32	2.74	2.37
$C_2H_6$	-3.49	-5.97	-6.47	-1.29	-1.54	-2.48	-2.98	2.20	1.95
$CH_3OH$	-4.69	-7.25	-7.63	-5.11	-4.78	-2.56	-2.94	-0.42	-0.09
$CH_3CCH$	-6.50	-8.94	-9.21	-4.70	-5.06	-2.44	-2.71	1.80	1.44
$C_3H_6^b$	-23.22	-32.10	-32.91	-14.62	-29.88	-8.88	-9.69	8.60	-6.65
$CH_3CHCH_2$	-5.09	-7.46	-7.90	-2.78	-3.08	-2.38	-2.81	2.31	2.01
$CH_3CHO$	-7.67	-9.54	-10.00	-4.20	-4.73	-1.87	-2.33	3.47	2.93
$C_3H_8$	-7.39	-12.30	-13.25	-3.14	-3.59	-4.91	-5.86	4.25	3.86
$C_2H_5OH$	-11.29	-16.36	-17.27	-9.40	-9.23	-5.06	-5.98	1.89	2.06
$H_2CCHCHCH_2$	-7.15	-10.04	-10.43	-5.02	-5.28	-2.90	-3.28	2.13	1.87
$C_4H_8^b$	-19.87	-29.98	-31.46	-10.05	-11.43	-10.11	-11.59	9.82	8.45
$(CH_3)_2CO$	-14.38	-17.69	-18.59	-7.78	-8.74	-3.31	-4.21	6.60	5.64
$CH_3COOH$	-15.84	-20.11	-21.14	-12.02	-12.13	-4.27	-5.29	3.82	3.71
$C_4H_4O$	-16.86	-25.37	-27.08	-13.30	-13.50	-8.31	-10.22	3.56	3.36
$C_6H_6$	-22.17	-33.24	-35.02	-15.90	-17.00	-11.08	-12.85	6.27	5.17
$C_6H_{12}$	-24.08	-29.75	-42.35	-11.34	-12.35	-15.67	-18.27	12.74	11.72

<sup>a</sup> The data are tabulated for the HF, MP2, QCISD(T), BLYP and B3PW91 methods. All the data are in kilocalories per mole.

<sup>b</sup> Data not plotted in Figures 1 and 2. See text.

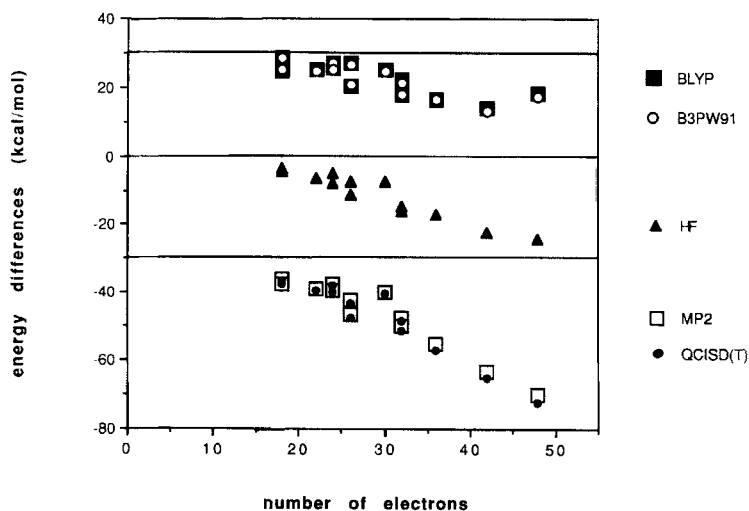
Plots of the data are given in Figures 1 and 2, where the various sets of data have been numerically displaced from one another for ease of presentation. The BLYP and B3PW91 approaches yielded similar results so they have been plotted together; the same is true for the MP2 and QCISD(T) calculations. The data have been arbitrarily plotted against the electron number of the molecule involved and show essentially a linear dependence on that variable. The linear regression parameters for the various plots are given in Table III along with the rmse and maximum error values. That data illustrate that calculation of energies using the locally dense basis set approach is viable and that, after correction by linear regression, the errors (of the order of 1.6–3.1 kcal/mol) clearly border chemical accuracy. The use of empirical correction factors in developing model chemistries is not unusual. For example, Pople and co-workers use such a factor based on the number of electrons in their G1 and G2 model chemistries<sup>21–25</sup>; the factor of 4.84 kcal/mol in the G2 method<sup>23,24</sup> is significantly larger than those used here.

The data shown in Figure 1 for the locally dense energy differences show that this approach overestimates the energy relative to that of the large balanced basis set calculation. As indicated in the figure caption, the various sets of data have been numerically displaced by +30, 0, and –30 kcal/mol for clarity of presentation; the horizontal

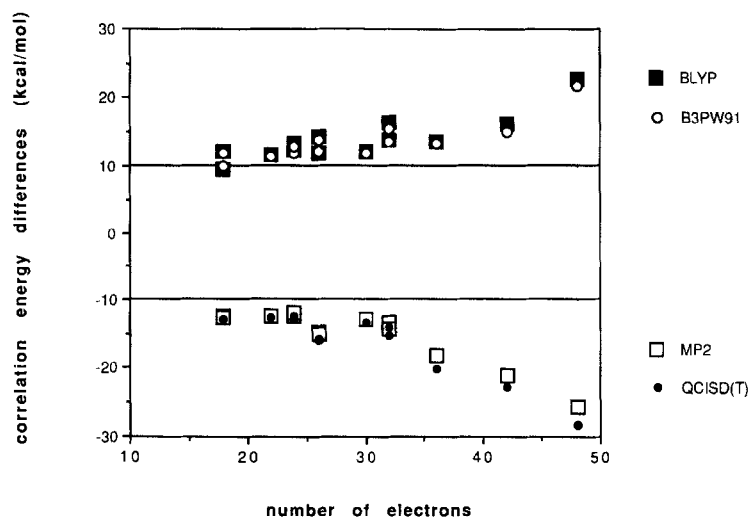
lines in the figure illustrate these numerical offsets and allow one to obtain a feeling for the errors involved by viewing each set of data relative to the nearby horizontal line. The use of the number of electrons as the plotting variable was chosen since the data were observed to relate simply in this manner and because exact correlation energies tend to vary in this way.<sup>53</sup> The observation that the plots are basically linear simply suggests that one is making a constant error per electron in the application of the locally dense basis set method.

Figure 2 shows the plot of correlation energy differences calculated in the locally dense basis set approach, the data again shifted (by +10 and –10 kcal/mol) for ease of presentation. As seen from the linear regression parameters in Table II, the variation of the fit with electron number is weaker than for the energy differences themselves, and the rmse values of 1.6 and 2.0 kcal/mol are clearly in the “chemical accuracy” range. The fact that the correlation energy differences are determined better than energy differences themselves suggests cancellation of error.

There is an especially interesting property of the density functional theory correlation energies as seen both in the locally dense and balanced basis set approaches. While the difference in correlation energy in the 6-311G(*d, p*) and 3-21G cases is large for a post-Hartree–Fock method like QCISD(T)



**FIGURE 1.** Differences between the locally dense and balanced energies (kilocalories per mole) for the HF, MP2, QCISD(T), BLYP, and B3PW91 methods shown in Table II plotted as a function of the number of electrons. The BLYP and B3PW91 data have been shifted up by +30 kcal/mol and that for the MP2 and QCISD(T) data have been shifted down by –30 kcal/mol for ease of presentation. The horizontal lines represent the shifted baselines for the various sets of data including the unshifted HF data.



**FIGURE 2.** Differences between the locally dense and balanced correlation energies (kilocalories per mole) for the MP2, QCISD(T), BLYP, and B3PW91 methods shown in Table II plotted as a function of the number of electrons. The BLYP and B3PW91 data have been shifted up by +10 kcal/mol and that for the MP2 and QCISD(T) data have been shifted down by -10 kcal/mol for ease of presentation. The horizontal lines represent the shifted baselines for the various sets of data.

varying from some 61 kcal/mol in acetylene to 244 kcal/mol in cyclohexane, the variation with basis set in the two density functional theory approaches is virtually nil. For BLYP the correlation energy differences between the big and small basis sets for the molecules studied show a mean of -3.6 kcal/mol with a standard deviation of 2.7 kcal/mol; the mean and standard deviation for the B3PW91 approach are +2.7 and 3.3 kcal/mol, respectively. Indeed, essentially the same constant values are found for the individual fragment calculations in each locally dense calculation. That is to say, the two density functional methods used

here show a surprising lack of sensitivity to basis set (balanced or mixed) in the determination of the correlation energy. This tends to suggest that the correlation energy is a relatively insensitive function of the electron density in density functional theory.

These observations regarding total energies and correlation energies are illustrated in Tables IV and V. Table IV contains the negative of the total energies for the HF calculations and the associated correlation energies for QCISD(T) as a representative post-Hartree-Fock calculation compared to the BLYP and B3PW91 density functional methods.

**TABLE III.**  
**Slope and Intercept Values for the Linear Regressions of the Locally Dense Energy and Correlation Energy Differences Relative to Balanced Basis Set Results.<sup>a</sup>**

	Intercept	Slope	rmse	Maximum Error
Energy differences				
HF	9.84 (2.1)	-0.726 (0.07)	1.96	4.80
MP2 + QCISD(T)	17.31 (2.2)	-1.178 (0.07)	3.06	7.98
BLYP + B3PW91	4.95 (1.8)	-0.431 (0.06)	2.56	4.39
Correlation energy differences				
MP2 + QCISD(T)	7.47 (1.4)	-0.452 (0.05)	1.97	4.07
BLYP + B3PW91	-4.89 (1.1)	0.295 (0.04)	1.56	3.45

<sup>a</sup> The intercept, root-mean-square error (rmse), and the maximum error are in kilocalories per mole and the slope is in kilocalories per mole per electron. The standard deviations of the regression parameters are shown in parentheses.



TABLE IV.

Negative Total Energies  $-E$  for the HF Calculations and Negative Correlation Energies  $-\Delta E_{\text{corr}}$  for the QCISD(T), BLYP, and B3PW91 Methods in the Balanced 6-311G(*d, p*) basis (energies are in atomic units).

	$-E$	$-\Delta E_{\text{corr}}$		
	HF	QCISD(T)	BLYP	B3PW91
C <sub>2</sub> H <sub>2</sub>	76.83914	0.30290	0.48461	0.47780
C <sub>2</sub> H <sub>4</sub>	78.05406	0.33300	0.51321	0.52590
C <sub>2</sub> H <sub>6</sub>	79.25160	0.36672	0.54153	0.57427
CH <sub>3</sub> OH	115.07493	0.39904	0.63987	0.63832
C <sub>3</sub> H <sub>4</sub>	115.89116	0.46985	0.74702	0.75107
CH <sub>3</sub> CCH	115.89391	0.46974	0.74006	0.74564
C <sub>3</sub> H <sub>6</sub>	117.08824	0.50553	0.76598	0.79988
CH <sub>3</sub> CHCH <sub>2</sub>	117.10269	0.50234	0.76861	0.79347
CH <sub>3</sub> CHO	152.95628	0.53664	0.87033	0.85911
C <sub>3</sub> H <sub>8</sub>	118.29566	0.53698	0.79545	0.84036
CH <sub>3</sub> CH <sub>2</sub> OH	154.12486	0.56861	0.89448	0.90460
CH <sub>2</sub> CHCHCH <sub>2</sub>	154.95862	0.63870	0.99833	1.01422
C <sub>4</sub> H <sub>8</sub>	156.13415	0.67851	1.01931	1.06668
(CH <sub>3</sub> ) <sub>2</sub> CO	192.01174	0.70551	1.12456	1.12575
CH <sub>3</sub> COOH	227.87551	0.73875	1.22516	1.19267
C <sub>4</sub> H <sub>4</sub> O	228.67852	0.81830	1.33129	1.31551
C <sub>6</sub> H <sub>6</sub>	230.75291	0.91180	1.44837	1.46315
C <sub>6</sub> H <sub>12</sub>	234.26310	1.01946	1.52331	1.59764

Recall that in all cases we define the correlation energy as the difference between the total energy calculated in any approach (post-Hartree-Fock or density functional theory) minus the energy of the associated Hartree-Fock calculation. The data in Table IV refer to the large basis set [6-311G(*d, p*)] basis used in the present work and show that for all the molecules investigated here the density functional theory correlation energies are consistently larger (more negative) than those for the QCISD(T) method by a not insignificant amount, covering a range of over 300 kcal/mol. This difference in correlation energy is illustrated in Figure 3 where the correlation energy per electron is plotted for the molecules shown in Table 4 and the three methods illustrated there. The correlation energy per electron for the density functional methods is approximately 21.3 kcal/mol while that for the QCISD(T) method is about 13.5 kcal/mol. These figures may be compared to the value of 17.1 kcal/mol which results from the atomic calculations of Chakravorty et al.<sup>53</sup> for the elements  $Z = 3-10$  plus that for helium as determined by Umrigar and Gonze.<sup>54</sup>

This difference in energy between the two density functional methods studied here and QCISD(T) is not limited to the 6-311G(*d, p*) basis set. Table V shows total energies and correlation energies for acetylene, ethylene, and ethane for

four different basis sets, including the very large 6-311++G(3*df*,2*p*) basis and estimations of the basis set limits for the methods tabulated. Feller<sup>55,56</sup> has demonstrated that the smooth convergence of the total energy of a molecule expanded in terms of the correlation consistent basis set family<sup>57,58</sup> can be fitted rather well with an exponential of the form

$$E = a + b \exp(-cn) \quad (7)$$

where  $n$  is the "size" of the basis set and the parameter  $a$  is the estimated basis set limit as  $n \rightarrow \infty$ . We have used the three largest basis sets indicated in Table V (the *i*, *b*, and *vb* sets) to obtain an exact fit to eq. (7) as a rough estimate of the basis set limit for the methods tabulated. An illustrative example of this type of fit is shown in Figure 4. While the "size" of a basis set with regard to energy convergence is difficult to define (Feller simply takes the index  $n$  as proportional to the basis set size), we have taken that parameter here to be the number of basis set functions for each molecule. The limiting correlation energy is again defined as the difference between the limiting total energy of a method and the limiting Hartree-Fock energy.

Two observations from the data in Table V are worth noting. First, as indicated previously, the density functional theory approaches consistently

TABLE V.

Negatives of the Total Energies and Correlation Energies for Acetylene, Ethylene, and Ethane as a Function of the Basis Sets 6-311++G(3df,2p) = vb, 6-311G(d,p) = b, 6-31G(d) = i, and 3-21G = s.<sup>a</sup>

	s	i	b	vb	Limit <sup>b</sup>
A. HCCH					
Energy					
HF	76.39399	76.81578	76.83914	76.84686	76.84690
MP2	76.57917	77.06678	77.11352	77.16020	77.16469
QCISD(T)	76.60014	77.09407	77.14203	77.18863	77.19283
BLYP	76.87252	77.29121	77.32374	77.33214	77.33216
B3PW91	76.87152	77.29053	77.31693	77.32401	77.32403
Correlation energy					
MP2	0.18519	0.25100	0.27439	0.31334	0.31779
QCISD(T)	0.20615	0.27830	0.30290	0.34178	0.34593
BLYP	0.47853	0.47544	0.48461	0.48528	0.48526
B3PW91	0.47754	0.47475	0.47780	0.47715	0.47713
B. H <sub>2</sub> CCH <sub>2</sub>					
Energy					
HF	77.60001	78.03098	78.05406	78.06240	78.06261
MP2	77.78123	78.28501	78.34730	78.39705	78.40399
QCISD(T)	77.81367	78.32211	78.38705	78.43694	78.44335
BLYP	77.11028	78.53679	78.56726	78.57540	78.57550
B3PW91	77.12936	78.55528	78.57995	78.78695	78.58705
Correlation energy					
MP2	0.18122	0.25403	0.29324	0.33464	0.34137
QCISD(T)	0.21366	0.29113	0.33300	0.37454	0.38073
BLYP	0.51027	0.50581	0.51321	0.51299	0.51288
B3PW91	0.52935	0.52430	0.52590	0.52455	0.52444
C. H <sub>3</sub> CCH <sub>3</sub>					
Energy					
HF	78.79355	79.22852	79.25160	79.25810	79.25829
MP2	78.98179	79.49474	79.57352	79.62412	79.63175
QCISD(T)	79.01665	79.53459	79.61832	79.66973	79.67682
BLYP	79.33271	79.76268	79.79312	79.79947	79.79957
B3PW91	79.37301	79.80170	79.82587	79.83157	79.83169
Correlation energy					
MP2	0.18824	0.26622	0.32192	0.36602	0.37346
QCISD(T)	0.22310	0.30606	0.36672	0.41163	0.41853
BLYP	0.53916	0.53416	0.54153	0.54137	0.54128
B3PW91	0.57946	0.57318	0.57427	0.57347	0.57340

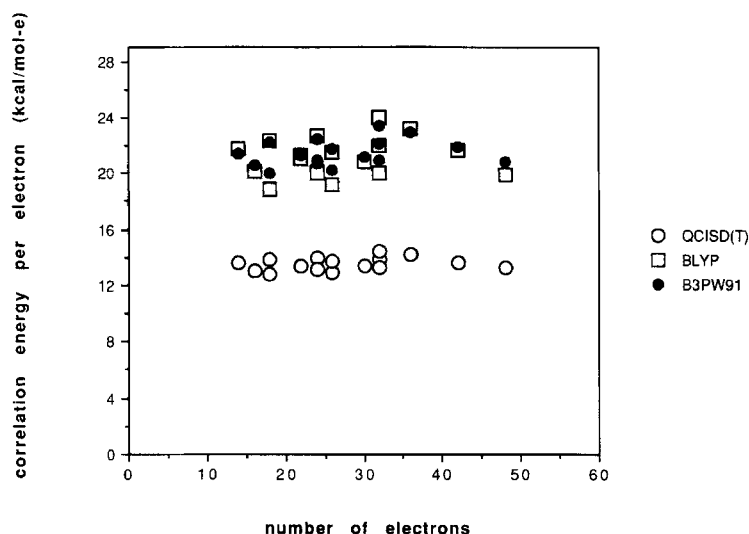
<sup>a</sup> Six *d* and ten *f* Cartesian functions were employed. All the energies are in atomic units.

<sup>b</sup> Limiting values are the estimated basis set limits found by Feller's exponential fit (refs. 55 and 56) to the three largest basis sets (vb, b, and i).

yield larger correlation energies than the post-Hartree-Fock methods, including the estimated basis set limits (being some 85–87 kcal/mol different). The other remarkable observation is the insensitivity of the correlation energies determined in the two density functional theory approaches as a function of basis set. While the spread of correlation energies over the three molecules is 88–123 kcal/mol for the QCISD(T) approach, its range is only 1.9–4.6 kcal/mol in the two density functional theory approaches.

#### VIABILITY OF THE LOCALLY DENSE BASIS SET APPROACH

Why should the locally dense set approach work? There is certainly no a priori reason to expect it to be successful. As previously noted, it tends to break symmetry and generally distorts the system's wave function from that obtained in a balanced basis set calculation. In the case of NMR chemical shielding, the rationale was offered<sup>2</sup> that the viability of the approach there results from the



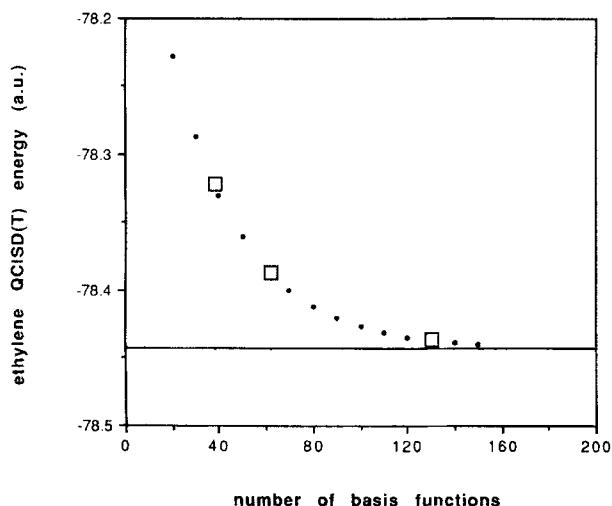
**FIGURE 3.** Correlation energy per electron (kilocalories per mole per electron) plotted against the number of electrons for the molecules shown in Table IV in the QCISD(T), BLYP, and B3PW91 methods using the balanced 6-311G(*d, p*) basis.

fact that NMR chemical shielding is a *local* effect, being most sensitive to the wave function in the vicinity of the resonant nucleus. However, we think of the energy as a global effect, and the fact that locally dense basis set calculations can reproduce total energies and correlation energies to a few kilocalories per mole is surprising. On the other hand, chemistry is replete with "localized" effects, including those involving energies. The CH bond

energy in one hydrocarbon is quite similar to that in others, the fact being attributed to the localized nature of chemical bonding.

We can frame an attempt to understand the behavior observed here in terms of localized molecular orbitals rather than the delocalized canonical molecular orbitals resulting from a Hartree-Fock calculation. Since the Hartree-Fock wave function is invariant to transformations which localize the molecular orbitals, a molecular system can be viewed as a set of localized systems of electrons interacting with other localized sets of electrons. The success of the locally dense basis set approach indicates that small modifications to the way we describe wave functions outside of a particular fragment have only a small effect on the energy contribution that particular fragment makes to the total molecular energy. This is not an explanation, but rather an observation that fits in with our ideas of the generally localized nature of chemical bonding.

While the present work has examined a representative set of molecules with two of the better known basis sets, only first-row species have been looked at and we have employed only one dense and one attenuated basis set. There is no reason to think that other covalent species should not fit the same pattern, but that remains to be tested. We did not look at any fluorine-containing molecules which might be a more stringent test of the method for first-row molecules. The loss of symmetry which accompanies the method only slightly re-



**FIGURE 4.** The QCISD(T) energy (in atomic units) for ethylene from Table V plotted against the number of basis functions (large open squares). The small solid dots represent the (exact) exponential fit of eq. (7) and the horizontal line is the estimated basis set limit.

duces its efficiency for post-Hartree-Fock methods, but, more importantly, does seem to affect more those molecules with orbital degeneracies near the HOMO-LUMO gap such as the strained cyclopropane and cyclobutane molecules. Practically speaking, the use of locally dense basis sets in density functional calculations is a failure, but their use there has revealed an interesting insensitivity of that method to basis set.

For post-Hartree-Fock methods, the locally dense basis set approach will allow one to investigate large chemical systems that would otherwise be inaccessible and to make reasonably accurate predictions of energy and correlation energy differences. The fact that the error in the methods is approximately linear in the electron number suggests that it may be applied without correction to systems where this quantity is conserved, such as the relative energy of isomers and chemical reactions themselves. It would be most interesting to see how the method behaves for open shell systems to see if bond dissociation energies might be accurately calculated. All in all, the use of locally dense basis sets to accurately estimate molecular energies appears to warrant further investigation.

## Acknowledgments

We are indebted to the North Carolina Supercomputing Center for an allocation of CPU time that made these calculations possible. E. F. C. B. was a participant in Duke University's 1995 NSF REU summer research program.

## References

1. D. B. Chesnut and K. D. Moore, *J. Comput. Chem.*, **10**, 648 (1989).
2. D. B. Chesnut, B. E. Rusiloski, K. D. Moore, and D. A. Egolf, *J. Comp. Chem.*, **14**, 1364 (1993).
3. C. Møller and M. S. Plesset, *Phys. Rev.*, **46**, 618 (1934).
4. M. Head-Gordon, J. A. Pople, and M. J. Frisch, *Chem. Phys. Lett.*, **153**, 503 (1988).
5. M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys. Lett.*, **166**, 275 (1990).
6. M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys. Lett.*, **166**, 281 (1990).
7. J. Cizek, *J. Chem. Phys.*, **45**, 4256 (1966).
8. J. Cizek, *Adv. Chem. Phys.*, **14**, 35 (1969).
9. G. D. Purvis and R. J. Bartlett, *J. Chem. Phys.*, **76**, 1910 (1982).
10. G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, III, *J. Chem. Phys.*, **89**, 7382 (1988).
11. G. E. Scuseria and H. F. Schaefer, III, *J. Chem. Phys.*, **90**, 3700 (1989).
12. J. A. Pople, M. Head-Gordon, and K. Raghavachari, *J. Chem. Phys.*, **87**, 5968 (1987).
13. J. A. Pople, M. Head-Gordon, and K. Raghavachari, *J. Chem. Phys.*, **87**, 5968 (1987).
14. K. A. Brueckner, *Phys. Rev.*, **96**, 508 (1954).
15. N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, *Chem. Phys. Lett.*, **164**, 185 (1989).
16. R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
17. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
18. M. L. McKee and W. N. Lipscomb, *J. Am. Chem. Soc.*, **103**, 4673 (1981).
19. J. V. Ortiz and W. N. Lipscomb, *Chem. Phys. Lett.*, **103**, 59 (1983).
20. R. H. Nobes, W. J. Bouma, and L. Radom, *Chem. Phys. Lett.*, **89**, 497 (1982).
21. J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.*, **90**, 5622 (1989).
22. L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.*, **93**, 2537 (1990).
23. L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.*, **94**, 7221 (1991).
24. L. A. Curtiss, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.*, **98**, 1293 (1993).
25. L. A. Curtiss, J. E. Carpenter, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.*, **96**, 9030 (1992).
26. S. Saebo and P. Pulay, *Ann. Rev. Phys. Chem.*, **44**, 213 (1993).
27. C. E. Dykstra and B. Kirtman, *Ann. Rev. Phys. Chem.*, **41**, 155 (1990).
28. D. B. Chesnut, In *Annual Reports on NMR Spectroscopy*, G. A. Webb, Ed., Academic Press, New York, 1994, Vol. 29.
29. D. B. Chesnut, In *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996.
30. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.*, **72**, 650 (1980).
31. M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.*, **80**, 3265 (1984).
32. J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980).
33. D. B. Chesnut and C. G. Phung, *Chem. Phys. Lett.*, **183**, 505 (1991).
34. D. B. Chesnut and C. G. Phung, In *Nuclear Magnetic Shieldings and Molecular Structure*, J. A. Tossell, Ed., NATO ASI Series, Kluwer Academic Publishers, Norwell, MA, 1993.
35. J. F. Hinton, P. Guthrie, P. Pulay, and K. Wolinski, *J. Am. Chem. Soc.*, **114**, 1604 (1992).
36. A. C. de Dios, J. G. Pearson, and E. Oldfield, *Science*, **260**, 1491 (1993).
37. H. Huber, *Chem. Phys. Lett.*, **112**, 133 (1984).
38. H. Huber and P. Diehl, *Mol. Phys.*, **54**, 725 (1985).
39. H. Huber, *J. Chem. Phys.*, **83**, 4591 (1985).
40. R. S. Grev and H. F. Schaefer, III, *J. Chem. Phys.*, **91**, 7305 (1989).

41. D. B. Chesnut and K. D. Moore, *J. Chem. Phys.*, **96**, 7188 (1992).
42. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Comperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian94, Revision A1, Gaussian Inc., Pittsburgh, PA, 1995.
43. R. J. Bartlett, *J. Phys. Chem.*, **93**, 1697 (1989).
44. K. Raghavachari, *Ann. Rev. Phys. Chem.*, **42**, 615 (1991).
45. A. D. Becke, *J. Chem. Phys.*, **98**, 5648 (1993).
46. A. D. Becke, *Phys. Rev. A*, **38**, 3098 (1988).
47. C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B*, **37**, 785 (1988).
48. P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, *Chem. Phys. Lett.*, **197**, 499 (1992).
49. P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, *Int. J. Quantum Chem. Symp.*, **26**, 319 (1992).
50. C. W. Murray, G. J. Laming, N. C. Handy, and R. D. Amos, *Chem. Phys. Lett.*, **199**, 551 (1992).
51. B. G. Johnson, P. M. W. Gill, and J. A. Pople, *J. Chem. Phys.*, **98**, 5612 (1993).
52. K. H. Hellwege, Ed., *Landolt-Börnstein, Zahlenwerte und Functionen, New Series II / 7*, Springer, Berlin, 1976.
53. S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. Fischer, *Phys. Rev. A*, **47**, 3649 (1993).
54. C. J. Umrigar and X. Gonze, *Phys. Rev. A*, **50**, 3827 (1994).
55. D. Feller, *J. Chem. Phys.*, **96**, 6104 (1992).
56. D. Feller, *J. Chem. Phys.*, **98**, 7059 (1993).
57. T. H. Dunning, Jr., *J. Chem. Phys.*, **90**, 1007 (1989).
58. R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *J. Chem. Phys.*, **96**, 6796 (1992).